
Réaliser un plan de gestion de données

Ce document a été conçu afin d'accompagner les chercheurs et chargés de projets lors de la rédaction de plans de gestion de données (*Data Management Plans, DMP*). Sa structure s'appuie sur le modèle proposé par la Commission européenne dans le cadre d'Horizon 2020 et divers modèles de plans de gestion de données existants tels que celui de la *National Science Foundation (NSF)* ou de l'*Interuniversity Consortium for Political and Social Research (ICPSR)*. Les champs requis par la Commission européenne sont signalés par un astérisque. Les exemples mentionnés dans ce document sont issus de guides existants.

Ce document constitue un guide de rédaction et non une liste de champs obligatoires.



Conception et réalisation : Aurore CARTIER (Service Commun de la Documentation, université Paris Descartes), Magalie MOYSAN (Bureau des archives, université Paris Diderot) et Nathalie REYMONET (Direction d'Appui à la Recherche et à l'Innovation (DARI), université Paris Diderot)

Les données de la recherche : définitions

Les données de la recherche recouvrent un périmètre large, dont la définition peut varier selon les approches des acteurs, les institutions, les financeurs ou encore les contextes nationaux, donnant lieu à une pluralité de définitions dont voici quelques exemples ci-dessous.

- **Définition de l'OCDE**

Les données de la recherche sont des enregistrements factuels (chiffres, textes, images et sons) utilisés comme sources principales pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires à la validation des résultats de recherche¹.

- **Définition de la section « Aurore »² de l'Association des archivistes français**

Les données de la recherche sont l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. A ce titre, elles constituent une partie des archives de la recherche.

- **Introduction des données de la recherche dans le Code de la recherche**

La loi sur l'enseignement supérieur et la recherche de juillet 2013 modifie l'article L 112-1 du Code de la recherche en introduisant la notion d'« organisation de l'accès libre aux données scientifiques »³.

- **Définition de la Commission européenne pour Horizon 2020**

1) *the data, including associated metadata (i.e. the metadata describing the research data deposited), needed to validate the results presented in scientific publications ;*

2) *other data (i.e. curated data not directly attributable to a publication, or raw data), including associated metadata⁴.*

Si les exigences de la Commission européenne sont à l'origine du présent outil, ce dernier a été conçu comme un canevas pouvant répondre aux exigences d'autres financeurs et à d'autres projets impliquant une gestion des données avec archivage. C'est la raison pour laquelle la définition des données de la recherche sur laquelle s'appuie ce travail a été volontairement élargie et s'attache davantage à la notion de « donnée » qu'à sa dimension « recherche » dont la définition peut varier selon la nature et le périmètre des projets concernés comme des acteurs impliqués.

En conséquence, le présent document considère les données de la recherche comme **un ensemble d'informations factuelles enregistrées sur des supports, produites ou collectées, selon divers procédés au cours d'un processus de recherche.**

¹ OCDE. *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* (2007)

² Section Aurore (Archives des universités, rectorats, organismes de recherche et mouvements étudiants) de l'Association des archivistes français, définition en cours de révision

³ Loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche, NOR: ESRJ1304228L

⁴ European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, v. 1.0, 11 décembre 2013, p. 14

Introduction au *Data Management Plan* (DMP)

Un *Data Management Plan* (DMP) ou plan de gestion de données est un document formalisé explicitant la manière dont seront obtenues, documentées, analysées, disséminées et utilisées les données produites au cours et à l'issue d'un processus ou d'un projet de recherche.

Le DMP s'appuie sur le cycle de vie des données/documents et décrit les choix réalisés en termes de normes de métadonnées⁵, formats des bases de données, méthodes et sécurité d'accès, durées d'archivage, ainsi que les coûts associés à la gestion des données. Une mention particulière doit être apportée aux données venant en appui des publications et qui doivent à ce titre rester disponibles et accessibles au plus grand nombre⁶.

L'établissement de *data management plans* est de plus en plus demandé dans les appels à projets financés sur fonds publics, notamment européens. Le guide de la Commission européenne sur l'*open access* des publications et des données dans Horizon 2020 en fournit la définition suivante :

*"A data management plan is a document outlining how the research data collected or generated will be handled during a research project, and after it is completed, describing what data will be collected/generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved."*⁷

L'objectif est ainsi de documenter la manière dont seront produites ou collectées les données au cours et à l'issue d'un processus de recherche, en s'attachant notamment à définir comment elles seront décrites, partagées, protégées puis conservées.

La vocation d'un DMP n'est pas de contraindre la pratique par l'exigence de nouveaux éléments mais plutôt de formaliser au sein d'un **document unique, prospectif, descriptif et évolutif**, un ensemble d'éléments et d'informations, auparavant dispersés entre divers acteurs et divers documents, éléments utiles au suivi du projet et à la bonne gestion des résultats obtenus.

⁵ Métadonnées : les métadonnées sont un ensemble de données structurées décrivant des ressources physiques ou numériques. Elles sont classiquement divisées en métadonnées descriptives, administratives ou de structure (<http://openaccess.inist.fr>).

⁶ CNRS, DIST. *Schéma d'orientation stratégique de l'information scientifique et technique* (Nov. 2013)

⁷ European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* (Dec. 2013), p. 10

Focus : Open Access et Data Management Plan dans les appels à projets Horizon 2020

L'*open access* des publications

L'*open access* des publications désigne le libre accès à l'information scientifique, défini comme un accès gratuit, immédiat, permanent, en ligne, aux articles publiés dans les revues à comité de lecture⁸.

Avec son programme cadre « Horizon 2020 », la Commission européenne fait de l'*open access* un principe général pour les publications scientifiques : elle rend obligatoire la publication en *open access* « en respect des intérêts légitimes des participants ». Il s'agit donc de rendre les publications scientifiques accessibles sans barrière financière, juridique ou technique. Les bénéficiaires s'engagent ainsi à respecter l'article 29.2 de la convention de subvention (*grant agreement*)⁹. Ils disposent toujours de la possibilité de publier leurs résultats de recherche auprès de revues à comité de lecture, mais le cas échéant, doivent le faire en *open access* selon différentes modalités : soit directement dans des revues en *open access* (*gold open access*) soit en déposant, parallèlement ou sous réserve d'embargo, leurs publications dans un serveur d'archives ouvertes (*green open access*)¹⁰.

L'*open access* des données

▪ Le périmètre de l'*Open research data pilot*

La Commission européenne lance un projet pilote intitulé "*Open research data pilot*" pour encourager et optimiser la gestion et la réutilisation des données de recherche générées par les projets qu'elle finance¹¹. L'utilisation d'un plan de gestion de données est obligatoire pour les projets participant au pilote, en dehors d'exceptions justifiées. Les autres projets de recherche peuvent intégrer le pilote s'ils le souhaitent. Une version initiale du plan de gestion de données doit être remise dans les six premiers mois suivant l'acceptation du projet. Le DMP peut alors être un document succinct d'une à deux pages mais devra être mis à jour au cours et à l'issue du projet.

Les responsables des projets financés par Horizon 2020 doivent prévoir la diffusion et la mise à disposition, en libre accès, des données de la recherche produites ou collectées dans le cadre de ces projets. Cette libre diffusion participe d'un cercle vertueux visant à améliorer la qualité

⁸ *Open access* : Budapest Open Access Initiative (2001), <http://www.opensocietyfoundations.org/openaccess>

⁹ *Each beneficiary must ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results.* Multi-beneficiary General Model Grant Agreement, Version 1.0 (11 December 2013)

¹⁰ European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, p.6

¹¹ European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, p. 8-11

des données, réduire la duplication des efforts de recherche, accélérer le progrès scientifique et contribuer à la lutte contre la fraude scientifique¹².

Le projet pilote recouvre les thématiques suivantes du programme Horizon 2020 :

Les 3 piliers du programme Horizon 2020 et le périmètre de l'Open research data pilot

EXCELLENCE SCIENTIFIQUE

Conseil européen de la recherche (ERC)

Actions Marie Skłodowska-Curie

Technologies futures et émergentes (FET)

Infrastructures de recherche :

- Développer de nouvelles infrastructures ...

- Intégrer et ouvrir les IR d'intérêt européen

- infrastructures électroniques

- Soutien à l'innovation...

PRIMAUTÉ INDUSTRIELLE

Technologie de l'information et de la communication (TIC)

Technologies clés génériques (KET)

Espace

Innovation dans les PME

Accès au financement à risque

DÉFIS SOCIÉTAUX

Santé, bien-être, vieillissement

Sécurité alimentaire, bioéconomie...

Energies sûres, propres, efficaces :

- Villes et communautés intelligentes

Transports intelligents, verts, intégrés

Climat, environnement, matières premières

Sociétés inclusives et novatrices et capables de réflexion

Sociétés sûres

Diffusion de l'excellence et élargissement de la participation

Science pour et avec la société

Institut européen d'innovation et technologie (IET)

Centre commun de recherche / Joint Research Center (JCR)

¹² Commission européenne. *Recommandation de la commission du 17.7.2012 relative à l'accès aux informations scientifiques et à leur conservation* (17.7.2012) C(2012) 4890 final

▪ Les conditions de l'*Open research data pilot*

Les bénéficiaires s'engagent à :

- **accepter l'article 29.3** ("*Open access to research data*") **de la convention de subvention** les engageant à déposer certaines des données (et métadonnées correspondantes) produites au cours du projet dans un réservoir ouvert de données (*research data repositories*) et à leur associer une licence libre permettant leur exploitation et leur réutilisation.¹³
- **produire dans les 6 premiers mois du projet de recherche un plan de gestion de données** qui est appelé à être vérifié, revu et complété tout au long du projet¹⁴ : des versions plus élaborées du DMP pourront être livrées au cours du projet, au moins à **mi-parcours** et au **rapport final**¹⁵.

Des exceptions à l'article 29.3 sont prévues, notamment si les données produites sont des données sensibles (données personnelles, données devant faire l'objet d'une exploitation industrielle, etc.). Ces exceptions sont négociées avec la Commission européenne au moment de la rédaction de la convention de subvention (*Grant agreement*).

Attention : les projets qui ne donnent pas lieu à diffusion des données doivent également comporter un plan de gestion. Le DMP concerne l'ensemble des données, y compris celles qui n'ont pas vocation à être diffusées.

Pour atteindre son objectif d'ouverture des données, la Commission européenne met **des moyens à la disposition des équipes de recherche** : éligibilité au remboursement des coûts¹⁶ ; mise à disposition d'un réservoir ouvert de données ([Zenodo](#)) ; outil numérique d'aide à la rédaction des plans de gestion (en cours de finalisation).¹⁷

¹³ European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, v. 1.0, 11 décembre 2013, p. 14 ; Commission européenne, *Guidelines on Data Management in Horizon 2020*, v. 1.0, 16 décembre 2013, p. 3

¹⁴ "Since DMPs are expected to mature during the project, more developed versions of the plan can be included as additional deliverables at later stages.", European Commission. *Guidelines on Data Management in Horizon 2020*, p. 3.

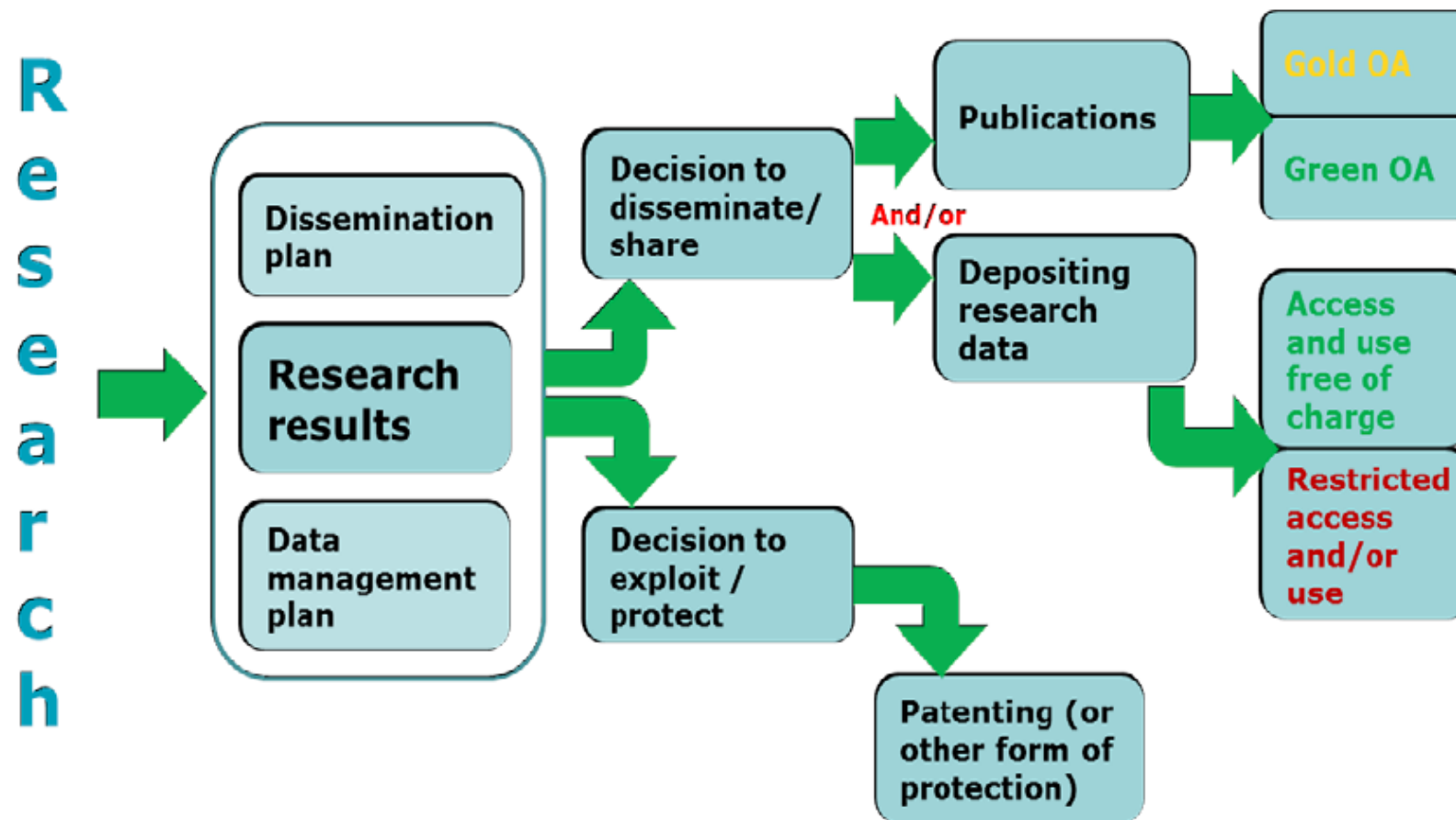
¹⁵ European Commission. *Guidelines on Data Management in Horizon 2020*, p. 3

¹⁶ "Costs relating to the implementation of the pilot will be eligible. Specific technical and professional support services will also be provided (e-Infrastructures WP)", European Commission. *Guidelines on Data Management in Horizon 2020*, p. 11

¹⁷ EUDAT. *Data access and management in the EUDAT collaborative data infrastructure (2014)*

Le schéma ci-dessous, produit par la Commission européenne, présente l'open access des publications et des données de la recherche dans le contexte de la dissémination et de l'exploitation des résultats¹⁸ :

OA in context: dissemination & exploitation of research results



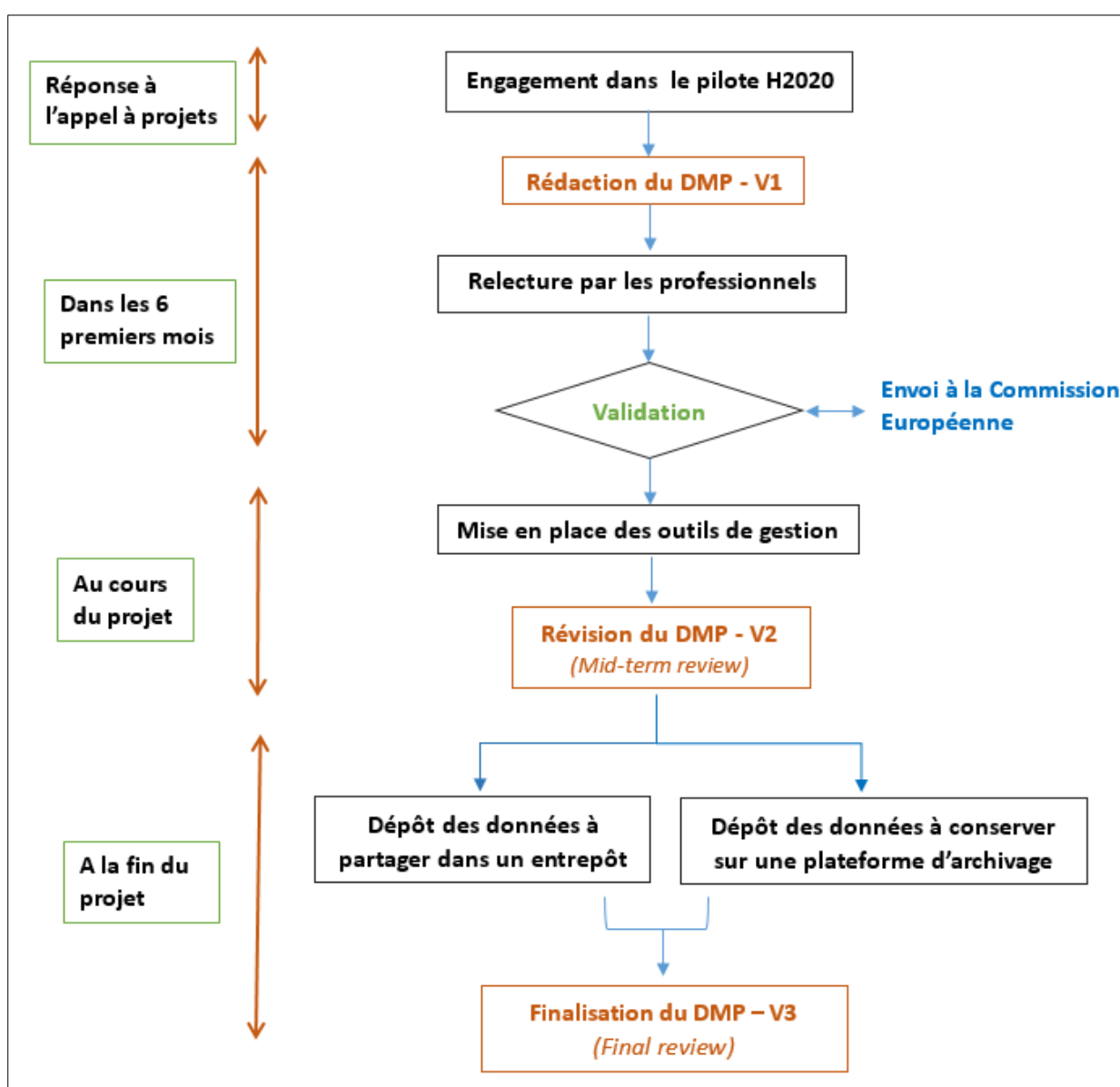
¹⁸ European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, n.4

Source: European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, p.4

Les étapes de rédaction et de validation du DMP dans le cadre d'Horizon 2020

Le DMP est un livrable requis par la Commission européenne. Il s'inscrit donc dans un calendrier précis. Sa rédaction et sa mise en œuvre reviennent à l'équipe de recherche lauréate, aidée des ingénieurs-projets européens et des professionnels de l'information scientifique et technique (IST).

Éléments de calendrier pour la réalisation d'un *Data Management Plan* dans H2020



Source : A. Cartier, M. Moysan, N. Reymonet, d'après les *Guidelines on Data Management in Horizon 2020*, de la Commission européenne.

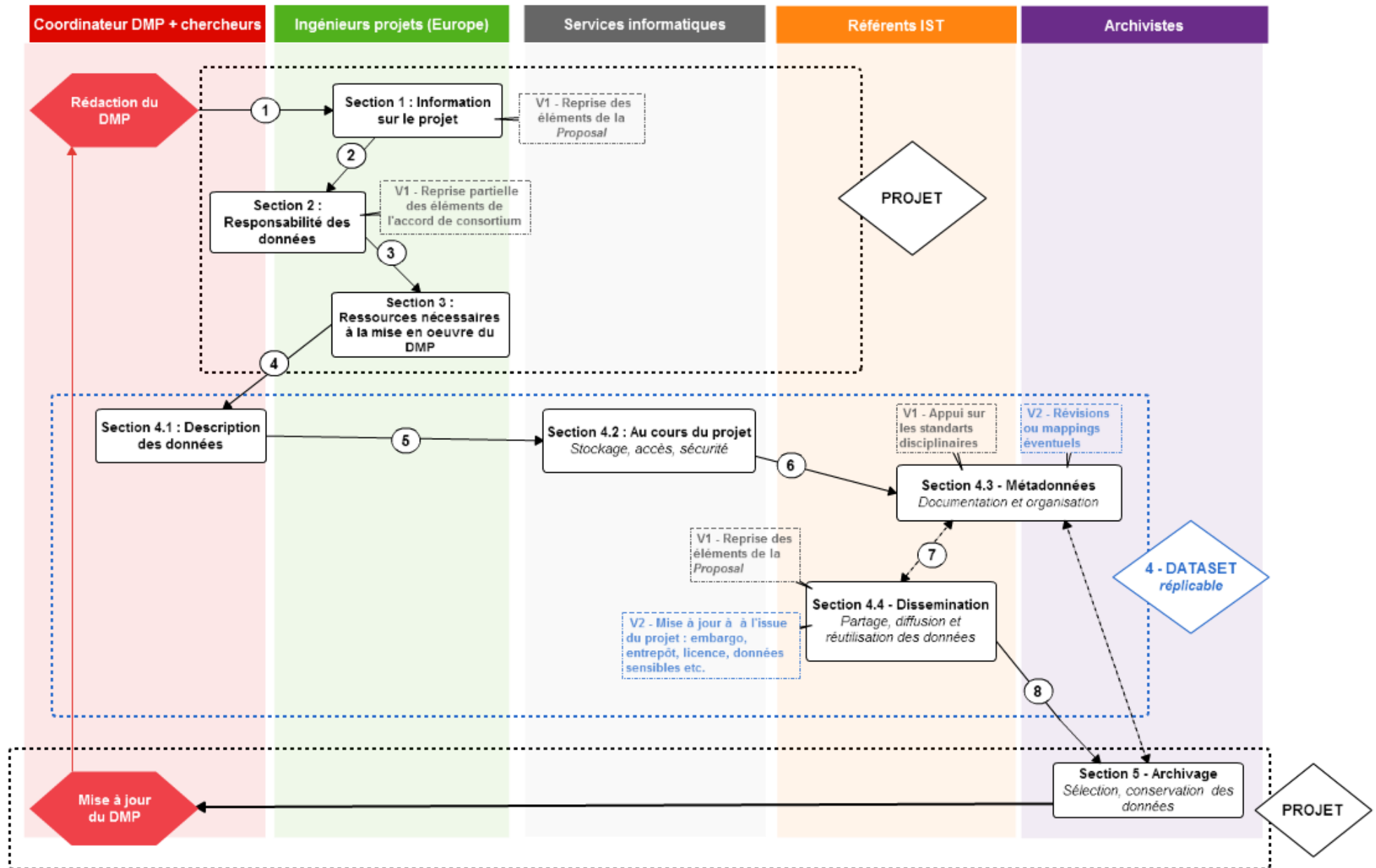
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

DMP : acteurs et services ressources

La rédaction de *data management plans* est une pratique née dans les pays anglo-saxons qu'il nous semble important d'adapter aux structures et acteurs français pour la rendre applicable. Si la rédaction d'un DMP relève davantage de la compilation et de la formalisation d'informations préexistantes que d'une démarche totalement inédite, force est de reconnaître qu'elle reste relativement coûteuse en temps pour les équipes de recherche. C'est pourquoi, à travers ce guide, notre démarche est non seulement de proposer une liste de champs applicables mais également d'identifier au sein de nos établissements d'enseignement supérieur et de recherche, les différents acteurs susceptibles d'accompagner les chercheurs dans la rédaction de leur DMP.

La rédaction d'un DMP fait en effet appel à diverses compétences bien souvent présentes au sein de nos universités mais généralement disséminées entre plusieurs acteurs, voire plusieurs services qui n'étaient pas, jusqu'ici, nécessairement habitués à travailler ensemble. Le schéma ci-dessous est une proposition de *workflow* pour la réalisation d'un *data management plan* appliqué au contexte des projets Horizon 2020 visant à indiquer, pour chaque étape du processus d'élaboration, les **personnes ressources** sur ces questions.

Les interlocuteurs ressources pour la rédaction et la mise en œuvre d'un plan de gestion de données en milieu universitaire



Références utilisées

- Agence des droits fondamentaux de l'Union européenne. *Manuel de droit européen en matière de protection des données*, 2014.
<http://fra.europa.eu/fr/publication/2014/manuel-de-droit-europeen-en-matiere-de-protection-des-donnees>
- Agence nationale de la sécurité des systèmes d'information. *Expression des besoins et identification des objectifs de sécurité (EBIOS®)*, http://www.securite-informatique.gouv.fr/gp_article82.html
- Budapest Open Access Initiative (2001) : <http://www.opensocietyfoundations.org/openaccess>
- CNRS, DIST. *Schéma d'orientation stratégique de l'information scientifique et technique* (nov. 2013)
<http://www.cnrs.fr/dist/docs/131119-orientation.pdf>
- Commission européenne. *Recommandation de la commission du 17.7.2012 relative à l'accès aux informations scientifiques et à leur conservation* (17.7.2012) C(2012) 4890 final
http://medoanet.sciencesconf.org/conference/medoanet/pages/recommandation_access_and_preservation_scientific_information_fr_copie.pdf
- Commission européenne /European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- Commission européenne /European Commission. *Guidelines on Data Management in Horizon 2020*
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- EDUCAUSE ACTI Data Management Working Group. *Developing an institutional research data management plan service* (jan. 2013) <https://net.educause.edu/ir/library/pdf/ACTI1301.pdf>
- EUDAT. *Data access and management in the EUDAT collaborative data infrastructure* (2014)
<http://www.eudat.eu/system/files/Open-Access-and-Data.pdf>
- ICPSR. *Guidelines for Effective Data Management Plans*
<http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf>
- Institut national de recherche agronomique. *Rapport du groupe de travail sur la gestion et le partage des données* (juin 2012) http://www.pfi-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf
- NSF. *Data Management Plan Requirements*. <http://www.nsf.gov/eng/general/dmp.jsp>
- OCDE. *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* (2007) <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>
- UC San Diego. *Data management plan* de Laurie Smith
http://rci.ucsd.edu/_files/DMP%20Example%20Laurie%20Smith.pdf
- UC San Diego. *Data management plan* d'Elsa Cleland
http://idi.ucsd.edu/_files/DMP%20Example%20Cleland.

CONTACTS

Coordination Sorbonne Paris Cité : Plan de gestion de données / *Data Management Plan*

- Aurore Cartier aurore.cartier@parisdescartes.fr
- Magalie Moysan magalie.moysan@univ-paris-diderot.fr
- Nathalie Reymonet nathalie.reymonet@univ-paris-diderot.fr

Plan de gestion de données / *Data management plan (DMP)*

Ce modèle de DMP, bilingue, est composé de plusieurs parties :

- **En-tête**
- **Section 1 : Informations sur le projet**
- **Section 2 : Responsabilité des données**
- **Section 3 : Ressources nécessaires à la mise en œuvre du DMP** : ces champs sont communs à l'ensemble du projet.
- **Section 4 : Jeux de données**
L'ensemble des sections ci-dessous doivent être dupliquées pour chaque jeu de données c'est-à-dire un ensemble de données techniquement homogène ou intellectuellement cohérent identifié comme tel.
 - **Section 4.1 : Description du jeu de données**
 - **Section 4.2 : Au cours du projet - Stockage, accès et sécurité des données**
 - **Section 4.3 : Métadonnées : documentation et organisation des données**
 - **Section 4.4 : A l'issue du projet : dissémination des jeux de données**
- **Section 5 : Sélection et archivage** se réalisent à l'échelle du projet, aussi cette partie concerne l'ensemble des données produites, qu'elles aient été diffusées ou non.

Les champs signalés par un astérisque* doivent être obligatoirement renseignés dans le cadre du pilote d'Horizon 2020

Nom du champ	Descriptif du champ	Exemples
0 – En-tête / 0 - Header		
Version du DMP, date <i>DMP version, date</i>		<i>Ex. : V2, 21 juillet 2014</i>
Date de la première version <i>Date of the first version</i>		<i>Ex. : 11 février 2014</i>

1 - Informations sur le projet / 1- Informations on the project		
<p>Cette section a pour vocation de renseigner administrativement sur le projet auquel ce DMP est lié, de le présenter et de le décrire succinctement.</p> <p><i>This section aims to provide administrative information about the project to which this DMP is linked and to present and briefly describe it.</i></p>		
Identifiant de l'appel à projets <i>Identification of the project call</i>		<i>Ex. : H2020-FETOPEN-2014-2015-RIA</i>
Thématique <i>Topic</i>		<i>Ex. : FETOPEN-1-2014</i>
Référence de la convention de financement <i>Grant agreement number</i>		<i>Ex. : 225260</i>
Programme de recherche <i>Research program</i>	Précise le programme de l'appel à projets.	<i>Ex. : Technologies futures et émergentes (FET)</i>
Acronyme du projet <i>Project acronym</i>		<i>Ex. 1 : TREASURE Ex. 2 : IPOD</i>

Titre du projet <i>Project title</i>		<i>Ex. : Etudes structurale et fonctionnelle des pompes à efflux OprN/MexE/MexF et OprM/MexC/MexD impliquées dans la résistance aux antibiotiques chez Pseudomonas aeruginosa</i>
Objectifs du projet <i>Goals of the project</i>	Description de la nature du projet, ses objectifs et son déroulement. Permet de comprendre le contexte et les types de données qui seront produites ou collectées au cours du projet.	<i>Ex. : l'objectif du projet est d'identifier les déterminants de la survie chez les patients porteurs du VIH et sous antirétroviraux.</i>
Mots-clefs du projet <i>Keywords</i>	Reprendre les mots-clefs de la proposition.	<i>Ex. : digital humanities, history, geography, ...</i>
Coordinateur/Bénéficiaire <i>Coordinator/Recipient</i>	Nom de l'institution, tel qu'il est connu par le financeur.	<i>Ex. 1 : Université Paris Diderot-Paris 7</i> <i>Ex. 2 : Université Paris Descartes</i>
Responsable scientifique du projet <i>Project leader</i>	Nom, prénom et mél.	<i>Ex. : Anne Leclerc</i> Anne.leclerc@univ-paris-diderot.fr
Affiliation et unité de rattachement du responsable scientifique <i>Administrative affiliation of the project leader</i>	Unité de rattachement du coordinateur du projet.	<i>Ex. 1 : Université Paris Diderot, UMR 7086</i> <i>Ex. 2 : Université Paris Descartes, U1153</i>

Nom du champ	Descriptif du champ	Exemples
2- Responsabilité des données /2 - Responsibility for the data		
<p>Cette section vise à identifier la ou les personne(s) qui seront en charge de la mise en œuvre et de la mise à jour du DMP tout au long du projet, ainsi que la propriété intellectuelle des données liées au projet de recherche.</p> <p><i>This section provides information about the person(s) who will be responsible for implementing and updating the DMP throughout the project, and the intellectual property of them.</i></p>		
<p>Nom du responsable de la gestion des données au cours du projet de recherche</p> <p><i>Name of the person in charge of data management during the project</i></p>	<p>Le responsable du plan de gestion de données est impliqué dans le pilotage du projet de sa création à son aboutissement. Il est responsable du plan de gestion, de sa rédaction, de son suivi et de son éventuelle modification au cours du projet de recherche. Il peut être différent du responsable scientifique.</p> <p>Dans le cas d'un projet associant des partenaires sur d'autres sites ou dans d'autres institutions, les responsabilités entre les partenaires doivent être précisées ici.</p>	<p><i>Ex. : Albert Dupont</i></p> <p>albert.dupond@univ-paris-diderot.fr</p>
<p>Propriété des données</p> <p><i>Data property</i></p>	<p>A qui appartiennent les données produites et/ou réutilisées ? Quels droits régissent vos bases de données et leurs contenus ? Comment se répartit la propriété intellectuelle sur les données ? etc.</p> <p>La propriété des données est fixée par l'accord de consortium définissant le rôle des partenaires du projet. La négociation des termes de l'accord en amont du projet et du DMP est essentielle car elle conditionne en partie les modalités de partage et de dissémination ultérieures des données.</p>	<p><i>Ex.1 : Résultats Propres : les Résultats seront la propriété du Partenaire qui a réalisé (ou fait réaliser par un sous-traitant, sans financement des autres Parties) les études/travaux qui sont à l'origine des Résultats. Le partenaire propriétaire des Résultats Propres sera seul juge de l'opportunité et du choix de leurs modes de protection et d'utilisation.</i></p> <p><i>Ex.2 : Résultats Communs : Si les Résultats ont été générés par les études/travaux de plus d'un Partenaire (ou développés par un sous-traitant, financé conjointement par deux ou plusieurs Partenaires), chaque Partenaire aura une part égale de copropriété sur ces Résultats.</i></p>

Nom du champ	Descriptif du champ	Exemples
3 - Ressources nécessaires à la mise en œuvre du DMP / 3 - Resources needed to implement the DMP		
<p>Estimez les compétences nécessaires à la mise en œuvre du DMP : gestion, curation (sélection, nettoyage, normalisation et enrichissement des données), conservation à long terme et les coûts associés.</p> <p><i>Estimate the skills needed to implement the DMP: management, curation (selection, cleaning, normalization and data enrichment), long-term conservation and associated costs.</i></p>		
Matériel Hardware	Estimez les espaces de stockage nécessaires au cours du projet et leur coût, ainsi que le coût de la plateforme d'archivage à l'issue du projet.	
Personnel Staff	Estimez le temps nécessaire à la gestion et la curation des données en ETP et le coût correspondant.	<i>Un ETP de [n]% a été budgété pour couvrir la gestion, la documentation et la curation des données.</i>
Formations Training	Estimez le coût des formations à la curation à prévoir.	
Montant financier* Costs	Estimez le montant global de la mise en œuvre du DMP.	<i>The [repository] has estimated their additional cost to archive the data is [insert euro amount]. This fee appears in the budget for this application as well.¹⁹</i>

¹⁹ ICPSR, *op. cit.* p. 10

Nom du champ	Descriptif du champ	Exemples
4 - Jeux de données / 4 - Datasets		
<p>Selon le projet, un jeu de données peut recouvrir des réalités différentes. En effet, un ou plusieurs jeu(x) de données peuvent être lié(s) au projet de recherche, et désigner : i) un lot techniquement homogène, ou ii) un lot intellectuellement cohérent même si celui-ci est composé de lots techniquement hétérogènes.</p> <p><i>Depending of the project, a dataset may cover different realities. In fact, a unique or multiple datasets may be linked to the research project and represent a technically uniform batch of data or an intellectually consistent one, potentially made of technically heterogeneous samples.</i></p>		
<p>Nombre de jeu(x) de données <i>Number of datasets</i></p>	<p>Cette partie précise le nombre de jeux de données qui seront produits et/ou reçus dans le cadre du projet.</p> <p><i>This part aims to present the number of datasets that will be produced and/or received in connection with the project.</i></p>	<p><i>Ex. : Datasets will be defined as all data generated by the proposed project</i></p> <p><i>Ex. : Datasets will be defined by each workpackage there are from</i></p>

Les sections 4.1 à 4.4 ci-dessous sont répliquables pour chaque jeu de données identifié pour le projet

Nom du champ	Descriptif du champ	Exemples
4.1 - Description des données – Jeu de données n°.../ 4.1 - Data description – Dataset #...		
Cette section a pour vocation de présenter le jeu de données qui sera produit et/ou reçu dans le cadre du projet. <i>This section aims to generally present the type of data that will be produced and/or received in connection with the project</i>		
Identifiant et nom du jeu de données* ²⁰ <i>Reference and name of data set</i>	Ce champ donne l'identification et/ou le nom du jeu de données décrit ci-après.	
Nature des données* <i>Nature of data</i>	Ce champ précise le type de données qui seront produites ou utilisées dans le cadre du projet : données d'enquête, échantillons, code logiciel, corpus de texte, photographies, etc.	<i>Ex. : Data and biological materials to be generated by the proposed project will mainly consist of images (e.g. confocal and transmission electron micrographs)</i> ²¹
Réutilisation de données existantes <i>Reuse of existing data</i>	Ce champ précise si le projet s'appuie sur la réutilisation de données ou jeux de données déjà existants et éventuellement produits par d'autres équipes de recherche.	<i>Ex. : Ce projet s'appuie sur des enquêtes qualitatives préexistantes, qui ont été numérisées et mises à disposition de la communauté.</i> <i>Ex. : Nous avons utilisé les données de 4 enquêtes cas-témoins réalisées par notre équipe depuis 1998.</i>
Méthode de production des données* <i>Method of production of data</i>	Ce champ vient préciser le processus de production ou d'élaboration des données utilisées : instrument de mesure, enquête, analyse, observation, compilation, simulation, etc. Indiquez s'il s'agit de données produites au cours du projet, de réutilisation de données préexistantes ou des deux. Indiquez les normes qualité utilisées le cas échéant.	<i>Ex. 1 : Les données ont été recueillies par questionnaire, puis intégrées dans une base de données Epidata, version 3.0.</i> <i>Ex. 2 : Les données ont été collectées suivant la norme ISO/CEI 17025 sur les « exigences générales concernant la compétence des laboratoires d'étalonnages et d'essais ».</i>

²⁰ L'astérisque signale les champs à remplir obligatoirement dans le cadre du programme Horizon 2020.

²¹ UC San Diego. *Data management plan* de Laurie Smith, p. 1 (Site de l'université de San Diego : <http://rci.ucsd.edu/dmp/examples.html>)

<p>Formats des données* (formats originels ou de conversion)</p> <p>Data standard</p>	<p>Indiquez les formats des données.</p> <p>Dans le cas où les données feraient l'objet de conversions, une simple mention du format d'origine est suffisante ; en revanche il est préférable de documenter a minima la correspondance (<i>mapping</i>) entre les formats. Une description détaillée des formats de conversion est obligatoire dans le contexte du DMP.</p> <p>Dans la mesure du possible, privilégiez des formats ouverts ou largement utilisés afin de faciliter le partage.</p>	<p><i>Ex. 1 : All data will be stored in digital form, either in the format in which it was originally generated (i.e. Metamorph files, for confocal images; Spectrum Mill files, for mass spectra with results of mass spectra analyses stored in Excel files; tiff files for gel images; Filemaker Pro files for genetics records), or will be converted into a digital form via scanning to create tiff or jpeg files (e.g. western blots or other types of results).²²</i></p> <p><i>Ex. 2 : Quantitative survey data files generated will be processed and submitted to the [repository] as SPSS system files with DDI XML documentation. The data will be distributed in several widely used formats, including ASCII, tab-delimited (for use with Excel), SAS, SPSS, and Stata. Documentation will be provided as PDF. Data will be stored as ASCII along with setup files for the statistical software packages. Documentation will be preserved using XML and PDF/A.²³</i></p>
---	--	---

²² UC San Diego. *Data management plan* de Laurie Smith, *op. cit.*

²³ ICPSR. *Guidelines for Effective Data Management Plans*, p. 6

Nom du champ	Descriptif du champ	Exemples
4.2 - Au cours du projet - Stockage, accès et sécurité des données /4.2 - During the project - Storage, access and security		
Cette section définit les modalités d'hébergement, de sauvegarde et d'accès aux données pendant la phase active du projet. <i>This section defines hosting, backup and data access during the active phase of the project.</i>		
Stockage et enregistrement des données <i>Storage and recording</i>		
Support des données <i>Medium of data</i>	Existe-t-il des données sur support papier ou autre support physique ?	<i>Ex. 1 : Les questionnaires remplis par les patients sont numérisés au fil de l'eau mais l'original est conservé sur support papier pendant la durée du projet.</i> <i>Ex. 2 : Les cahiers de laboratoire sont conservés sur support papier.</i>
Volumétrie prévisionnelle* <i>Projected volume</i>	En mesure de stockage. Elle pourra être réévaluée au cours du projet.	<i>Ex. : [n] Megaoctets</i>
Type d'hébergement <i>Data hosting</i>	Décrivez le matériel et le lieu physique de stockage. Avez-vous recours aux services d'un prestataire pour le stockage de vos données au cours de la recherche ? A un logiciel spécifique ?	<i>Ex. : Serveur local, serveur distant, disque dur externe, etc.</i>
Sécurité des données <i>Data security</i>		
Risques ou menaces sur les données <i>Risks or threats to data</i>	Indiquez les risques susceptibles de menacer les exigences de confidentialité, de disponibilité et d'intégrité des données tout au long de la phase active du projet. Un référentiel de menaces est disponible dans la méthode EBIOS. ²⁴	<i>Ex. : Destruction accidentelle/illicite de données, altération, accès non autorisé, environnement à risque, etc.</i>

²⁴ EBIOS : Expression des besoins et identification des objectifs de sécurité. C'est une démarche d'analyse de sécurité élaborée par la Direction centrale de la sécurité de systèmes d'information du Secrétariat de la défense nationale (SGDN). Pour davantage d'informations, voir le portail de la Sécurité informatique de l'Agence nationale de la sécurité des systèmes d'information : http://www.securite-informatique.gouv.fr/gp_article82.html

<p>Garantie de confidentialité des données <i>Data privacy</i></p>	<p>Préciser les précautions prises pour garantir la confidentialité des données au regard des éventuels risques détectés.</p> <p>Selon la norme ISO 7498-2 (ISO90), la confidentialité est la propriété d'une information qui n'est <i>ni disponible ni divulguée</i> aux personnes, entités ou processus non autorisés. Dans le droit européen, tous les responsables de traitement ou sous-traitants doivent veiller à préserver la confidentialité des données²⁵.</p>	<p><i>Ex. : Une accréditation particulière sera décernée à toutes les personnes susceptibles d'accéder aux données.</i></p>
<p>Garantie d'intégrité et de traçabilité <i>Data integrity and traceability</i></p>	<p>Préciser les précautions prises pour garantir l'intégrité et la traçabilité des données au regard des éventuels risques détectés.</p> <p>Selon la norme ISO 7498-2 (ISO90), l'intégrité est la <i>prévention d'une modification</i> non autorisée de l'information. Précise les outils et précautions mis en œuvre pour suivre la production et l'analyse des données au cours du projet.</p>	<p><i>Ex. : Cahiers de laboratoire²⁶, protocoles de recherche, registres de consignation, horodatage des données et résultats, authentification personnelle sur les outils, etc.</i></p>
<p>Accès aux données <i>Access to data</i></p>		
<p>Lecture des données* <i>Data reading</i></p>	<p>La lecture des données nécessite-t-elle le recours à un logiciel ou un outil spécifique ? Si oui, lequel ?</p>	
<p>Garantie de disponibilité des données <i>Data availability</i></p>	<p>La disponibilité est la propriété d'<i>accessibilité</i> au moment voulu des données et des fonctions par les utilisateurs autorisés.</p>	
<p>Gestion des accès* <i>Access procedures</i></p>	<p>De quelle manière l'ensemble des chercheurs partenaires du projet auront-ils accès aux données pendant la recherche ?</p>	<p><i>Ex. : The data will be processed and managed in a secure non-networked environment using virtual desktop technology.²⁷</i></p>
<p>Echanges et partage* <i>Data sharing</i></p>	<p>Les données feront-elles l'objet d'échange ou de partage avec de tiers acteurs ?</p>	<p><i>Ex. : Les données collectées seront partagées avec le Groupe de recherche [...] de l'Université X... en charge de la thématique Y.</i></p>

²⁵ Agence des droits fondamentaux de l'Union Européenne, *Manuel de droit européen en matière de protection des données*, 2014, p. 101

²⁶ <http://www.inpi.fr/fr/vous-etes/laboratoire-ou-chercheur/le-cahier-de-laboratoire.html>

²⁷ ICPSR, *op. cit.* p. 7

Nom du champ	Descriptif du champ	Exemples
4.3 - Métadonnées : documentation et organisation des données / 4.3 - Metadata : documentation and data organization		
<p>Cette section précise la manière dont seront décrites et organisées les données produites ou reçues au cours du projet. <i>This section details how the data generated or received during the project will be described and organized.</i></p>		
<p>Standards et formats disciplinaires des métadonnées* Standards and metadata</p>	<p>Citez les standards et formats utilisés.</p> <p>Certaines disciplines s'appuient sur des standards et des formats de métadonnées qui leur sont propres. Ils doivent être utilisés en priorité pour assurer l'interopérabilité des données entre les différents systèmes et les différents utilisateurs.</p> <p>Pour s'informer sur les formats disciplinaires : http://www.dcc.ac.uk/resources/metadata-standards</p>	<p><i>Ex. 1: Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively.²⁸</i></p> <p><i>Ex. 2: The clinical data collected from this project will be documented using CDISC metadata standards.²⁹</i></p> <p><i>Ex. 3 : Ecological Metadata Languages (EML)</i></p> <p><i>Ex. 4 : Crystallographic Information Framework (CIF)</i></p>
<p>Mode de production et responsabilité des métadonnées Method of production and metadata responsibility</p>	<p>Indiquez par qui les métadonnées seront renseignées, de quelle manière et éventuellement à l'aide de quels outils.</p>	<p><i>Ex.: All datasets will be annotated with metadata. As data are generated they will be entered into Morpho, a free resource for associating Ecological Metadata Language (EML) with archived datasets. It will be the responsibility of each researcher to annotate their data with metadata, and it will be the responsibility of the PIs to check weekly (during the field season, monthly otherwise) with all participants to assure data is being properly processed, documented, and stored.³⁰</i></p>

²⁸ ICPSR. *Guidelines for Effective Data Management Plans*, p. 4

²⁹ ICPSR, *op. cit.* p. 4

³⁰ UC San Diego. *Data management plan* d'Elsa Cleland <http://idi.ucsd.edu/files/DMP%20Example%20Cleland.pdf>

Arborescence de classement <i>Tree classification</i>	Prévoyez une arborescence de classement des données dès que possible afin de faciliter l'accès et le stockage au cours du projet, puis à terme l'archivage des données.	<i>Ex. : Arborescence en cours de validation.</i>
Règles de nommage des jeux de données <i>Rules for naming data sets</i>	Précisez les règles de nommage envisagées pour harmoniser le nom des jeux de données.	<i>Ex. : sujet_typologie_annee_mois_jour_version</i>
Documentation associée <i>Relevant documentation</i>	Ce champ vise à recueillir toute autre information nécessaire à l'intelligibilité à long terme des données : codes, abréviations, version des logiciels de lecture, etc.	

Nom du champ	Descriptif du champ	Exemples
4.4 - A l'issue du projet – Dissémination / 4.4 - At the end of the project – Dissemination		
<p>Partage, diffusion et réutilisation des données / Data sharing, diffusion and reuse</p> <p>Cette section précise les modalités et les éventuelles précautions éthiques, juridiques et techniques selon lesquelles seront diffusées les données.</p> <p><i>This section describes the procedures and specifies any ethical, legal and technical safeguards under which the data will be released.</i></p>		
<p>Principe général de diffusion * <i>General principle of diffusion</i></p>	<p>Précise les règles de diffusion qui s'appliquent à la majorité des données produites dans le cadre du projet, selon les droits associés.</p>	<p><i>Ex. 1 : H2020. En vertu de l'article 29.3 des conventions de financement du 8^e PCRD (H2020), la majorité des données et métadonnées nécessaires à la validation des résultats doivent être déposées dans des entrepôts dédiés et placées sous une licence libre autorisant leur exploitation et leur libre réutilisation par une tierce partie.</i></p> <p><i>Ex. 2: ERC. "Annex 1 sets out how research data will be handled..."</i></p>
<p>Type de licence <i>Type of license</i></p>	<p>Précisez les conditions de partage et de réutilisation des données diffusées dans le cadre du projet ainsi que la licence et les éventuelles contreparties économiques qui peuvent y être associées.</p>	<p><i>Ex. 1 : L'ensemble des données diffusées seront placées sous licence libre Etalab.</i></p> <p><i>Ex. 2 : Les bases de données seront placées sous licence GNU</i></p>
<p>Potentiel de réutilisation * <i>Potentiel for reuse</i></p>	<p>Public cible envisagé (grand public, communauté scientifique, secteur privé, etc.) et perspectives d'application ou de développement à partir de vos données.</p>	<p><i>Ex. 1 : Les données climatologiques peuvent servir de sources pour des travaux de recherche en histoire ou en géographie.</i></p> <p><i>Ex. 2: In addition to the research community, we expect these data will be used by practioners and policymakers.³¹</i></p>

³¹ ICPSR, *op. cit.* p. 9

<p>Existence de publications associées aux données <i>Existing publications related to the data</i></p>	<p>Ce champ sera à renseigner à la fin du projet.</p>	<p><i>Oui/non</i></p>
<p>Dépôt et dissémination des données* <i>Data repository and access</i></p>	<p>Existe-t-il une base de données disciplinaire en ligne dans laquelle ces données pourraient être déposées ? Sinon, dans le cadre d'Horizon 2020, le dépôt de données peut se faire dans Zenodo³².</p>	<p><i>Ex. 1 : Les données de recherche issues de ce projet seront déposées dans Zenodo afin d'assurer leur accès à long terme par la communauté scientifique.</i></p> <p><i>Ex. 2: The project team will create a dedicated Web site to manage and distribute the data because the audience for the data is small and has a tradition of interacting as a community. The site will be established using a content management system like Drupal or Joomla so that data users can participate in adding site content over time, making the site self-sustaining. The site will be available at a .org location.³³</i></p>
<p>Protection des données sensibles / <i>Protection of sensitive data</i></p> <p>Pour diverses raisons éthiques, juridiques, financières ou encore techniques, certaines données peuvent nécessiter une protection spécifique et à ce titre échapper aux principes de diffusion. Cette section a donc pour but d'identifier et de définir les critères de protection des données sensibles susceptibles d'être produites ou collectées dans le cadre du projet.</p> <p><i>For a variety of ethical, legal, financial or technical reasons, data may require specific protection and as such, escape the rules of distribution. This section identifies and defines criteria for protection of sensitive data that can be produced or collected as part of the project.</i></p>		
<p>Identification des jeux de données sensibles <i>Identification of sensitive data sets</i></p>	<p>Ce champ répertorie les jeux de données susceptibles d'échapper au principe de diffusion : données scientifiques protégées ou à risque, données de gestion, données nominatives, données stratégiques...</p>	<p><i>Ex.: This project will generate data designed to study the prevalence and correlates of DSM III-R psychiatric disorders and patterns and correlates of service utilization for these disorders in a nationally representative sample of over 8000 respondents. The sensitive nature of these data</i></p>

³² <http://zenodo.org/>

³³ ICPSR, *op. cit.* p. 3

		<i>will require that the data be released through a restricted use contract.³⁴</i>
Justification du principe d'exception aux conditions générales de diffusion* <i>Justification for the exception to the general principles of diffusion</i>	<p>Ce champ détaille la nature du risque ou de la contrainte pesant sur les données susceptibles de justifier l'exception au principe de diffusion. Le risque peut être de nature variée : exposition de ressources rares (biodiversité, archéologie, etc.), risque industriel, respect de la vie privée, priorité d'exploitation, etc.</p>	<p><i>Ex. 1 : Données à caractère personnel.³⁵</i> <i>Ex. 2 : Secret défense (Cf. instruction et recommandations ministérielles du SGDN).</i> <i>Ex. 3 : Demande de brevet en cours.</i></p>
Mesures de protection <i>Precautionary measures</i>	<p>Ce champ détaille la façon dont la protection sera mise en œuvre.</p>	<p><i>Ex. : Les données à caractère personnel seront anonymisées avant partage et diffusion selon les recommandations de la CNIL.</i></p>
Embargo* <i>Embargo period</i>	<p>L'embargo est une protection temporaire qui permet de remettre à une certaine date la libre diffusion et/ou la réutilisation des données. L'embargo ne nécessite pas d'autorisation autre que celle du propriétaire. Les conditions de secret ou de discrétion au cours d'une période d'embargo sont moindres qu'une mesure de confidentialité.</p>	<p><i>Ex. 1 : Les données seront diffusées après une période d'embargo de 6 mois.</i> <i>Ex. 2 : Les données astronomiques sont destinées à être diffusées mais bénéficient d'une durée d'embargo d'un an pour priorité d'exploitation.</i></p>

³⁴ ICPSR, *op. cit.* p. 2

³⁵ Les traitements de données à caractère personnel doivent faire préalablement l'objet d'une déclaration à la Commission nationale informatique et libertés (CNIL). <http://www.cnil.fr/vos-obligations/declarer-a-la-cnil/>

Nom du champ	Descriptif du champ	Exemples
5 - Sélection et archivage des données / 5 - Data selection and long term preservation		
<p>Cette section concerne l'ensemble des données produites ou collectées au cours du projet, qu'elles aient été diffusées ou non. Il est fortement recommandé de prendre contact avec l'archiviste de votre établissement lors de la rédaction de cette section.</p> <p><i>This section applies to all data generated or collected during the project, whether distributed or not. It is strongly recommended to contact the archivist of your establishment to draft this section.</i></p>		
Sort des données à l'issue du projet <i>Fate of data at the end of the project</i>	Indique le devenir des données au terme de la phase active du projet et de la dissémination et valorisation des résultats.	<i>Ex. : A l'issue du projet, les données à caractère personnel seront détruites conformément aux recommandations de la CNIL. Un échantillon sera archivé pour mémoire. Les autres données seront archivées sur une plateforme d'archivage pérenne. Une copie sera conservée sur un disque dur par le responsable du projet.</i>
Sélection des données* <i>Data selection</i>	Les données n'ont pas toutes vocations à être archivées sur le long terme, néanmoins le visa de l'archiviste est recommandé avant toute opération de destruction massive de données. Indiquez les jeux de données qui auront vocation à être conservés à moyen ou long terme en raison de leur valeur scientifique, juridique, ou patrimoniale. En cas de doute, prenez contact avec l'archiviste de votre structure.	<i>Ex. 1 : Les données climatiques ne sont pas reproductibles et ont un intérêt à être conservées.</i> <i>Ex. 2 : Les données astronomiques ont vocation à être conservées car leur coût de production est extrêmement élevé.</i>
Volume final des données* <i>Final volume of data</i>	Indiquer le volume final estimé.	<i>Ex. : Elles représentent [n] Giga-octets.</i>
Durées de conservation préconisée* <i>Recommended lifetime</i>	Cette durée varie fortement. Elle peut aller de quelques mois à l'éternité. Les durées de conservation doivent prendre en compte les exigences légales et réglementaires existantes.	<i>Ex. : décret 2006-6 du 4 janvier 2006 fixant la conservation des données de patients à 20 ans minimum.</i>

<p>Plateforme d'archivage* <i>Long term preservation platform</i></p>	<p>Sur quelle plateforme d'archivage pérenne ces données seront-elles archivées à long terme ?</p> <p>Les plateformes d'archivage électronique doivent être agréées par le Ministère de la Culture. Certaines universités disposent de conventions avec des plateformes : contactez le service d'archives de votre institution.</p>	<p><i>Ex. : les données seront archivées dans le centre informatique [X].</i></p>
---	---	---